IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON ARTIFICIAL INTELLIGENCE FOR PHYSICAL-LAYER WIRELESS COMMUNICATIONS

Received June 11, 2019, accepted July 3, 2019, date of publication July 9, 2019, date of current version July 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2927644

Intelligent Contactless Gesture Recognition Using WLAN Physical Layer Information

SHUJIE REN[1,2], HUAIBIN WANG[1,2], LIANGYI GONG[1,2,3], CHAOCAN XIANG[4], XUANGOU WU[5], AND YUFENG DU[1,2]

[1]Key Laboratory of Computer Vision and System, Ministry of Education, Tianjin University of Technology, Tianjin 300384, China
[2]Tianjin Key Laboratory of Intelligence Computing and Novel Software Technology, Ministry of Education, Tianjin University of Technology, Tianjin 300384, China
[3]School of Software and BNRist, Tsinghua University, Beijing 100084, China
[4]College of Computer Science, Chongqing University, Chongqing 400044, China
[5]School of Computer Science and Technology, University of Anhui Technology, Maanshan 243002, China

Corresponding author: Liangyi Gong (gongliangyi@gmail.com)

This work was supported in part by the Natural Science Foundation of Tianjin Projects under Grant 18JCQNJC69900, in part by the NSF China Projects under Grant 61872447, Grant 61672038, and Grant 6160206, and in part by the Natural Science Foundation of Chongqing under Grant CSTC2018JCYJA1879.

ABSTRACT Contactless gesture recognition is an emerging interactive technique in ubiquitous and mobile computing. It combines the linguistics with the wireless signals to analyze, judge, and integrate human gestures by the usage of intelligent algorithms. The existing contactless gesture recognition studies can achieve gesture recognition with the machine learning technologies. But in practice, some objective factors, such as the user's position, the non-line of sight condition, can seriously affect the performance of these gesture recognition systems. In this paper, we propose an intelligent and robust contactless gesture recognition using physical layer information. Instead of the usage of machine learning, we learn the gesture characteristics based on the Fresnel zone model of wireless signals. First, we denoise the collected channel state information (CSI) in a sliding window. Then, we extract the eigenvalues of channel phase information based on Fresnel zone model to depict four basic gestures. The features of gestures are independent of the user's position and the signal amplitude. Finally, common-gesture recognition is achieved based on the decision tree classification. Moreover, we develop a hidden Markov model to achieve the complex-gesture recognition. The extensive experimental results show that our proposed method is position-independent and robust. The accuracy of basic-gesture recognition is as high as 91% on average. And, the accuracy of the complex-gesture recognition is also above 85% on average.

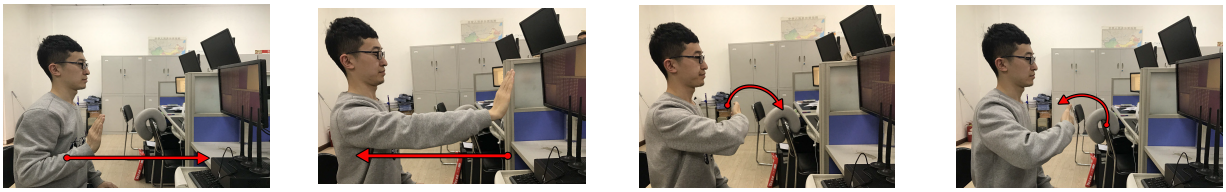
INDEX TERMS Contactless, Fresnel zone, channel state information, gesture recognition.

I. INTRODUCTION

With the rapid development of ubiquitous and mobile computing, more users wish to interact with smart devices in a contactless way. Traditional interaction with wearable devices requires the physical touch of the user, such as a mouse, a keyboard, or other peripheral devices, which causes much inconvenience for daily life and work. Compared with traditional ones, gesture-based interaction (remote recognition of gestures) can provide a more convenient and natural way for users to interact with smart devices. The intelligent applications, like smart home, virtual reality and body sense games, can learn the commands by the gestures, as shown in Fig. 1. For example, a user can control a home lamp by doing a slip-hand movement in the air without touching household goods. Or you may control the size of the sound on the sofa. In a formal presentation, users can also control powerpoint page flipping without touching the computer but by sliding hands in the air.

Traditional contactless gesture recognition can be divided into two types: (1) one based on professional hardware equipment. The camera-based gesture recognition systems can identify the gestures of users by images [1], [2], nevertheless only working under the conditions of line-of-sight path. Besides, users can also wear specialized sensors, such as smart wristbands [3], hand-held sensors [4], [5]. These

The associate editor coordinating the review of this manuscript and approving it for publication was Guan Gui.



(a) Push diagram (b) Pull diagram (c) Right diagram (d) Left diagram

FIGURE 1. Four different gestures defined in our experiments.

wear-based gesture recognition methods solve the challenge of an invalid of non-line-of-sigh. However, wearable devices bring inconvenience due to the limited power, even for the older people. (2) the other based on radio frequency equipment. This method does not require the user to wear any physical sensor. It can also work under non-line-of-sight conditions [6]–[11]. However, the system is realized based on machine learning techniques. It requires extensive training and learning in the early stage, and has a high limitation on the environment where the user is located.

In order to overcoming existing challenges in contactless gesture recognition, we propose an intelligent and robust WiFi-based contractless gesture recognition, called iGest. iGest uses the radio frequency sensing technique for position-independent gesture recognition in the indoor environment. It can be used to recognize the gestures in a contactless way with the following advantages. (1) In terms of equipment, we use commercial WiFi cards to achieve gesture recognition. (2) We extract the time-domain features of gestures based on the Fresnel zone model. The features are not related to the signal amplitudes and independent of the user's positions.

Then, we design and develop the position-independent gesture recognition system based on the commercial WiFi cards. First, we make a data processing for the physical layer information of WiFi signals, including data collecting, noise removal and window slice. The pure channel phase information can be selected for the next feature extraction. Next, we utilize the feature extraction to describe the different gestures. The phase difference between subcarriers can be extracted and used to calculate peak and valley features. We encode eigenvalues after extracting the peak and valley values. Finally, we use features coding for gesture recognition based on the decision tree classification. Moreover, a hidden Markov model is proposed to recognize the complex gestures.

The prototype system of iGest is designed and tested in three environments. We use two mini computers as the receiver and the transmitter respectively, while selecting the physical layer information of WiFi from the Intel 5300 NIC. In our experiments, we employ four volunteers to perform four basic gestures and four complex gestures. We totally take about 160 groups of gestures recognition in different locations of different environments. The experimental evaluation shows that the iGest can achieve good performance. It can overcome the limitation on the position of the human body. Gestures can also be recognized in a non-line-of-sight environment. In addition, no special sensing equipment is required. Moreover, the system overcomes the shortcomings of coarse granularity recognition based on the RF device. And it can recognize human gestures with fine granularity.

In summary, the main contributions of this work are as follows:

- We propose iGest an intelligent and robust contactless gesture recognition using WiFi physical layer information. It is a fine-grained and location-independent gesture recognition system.
- iGest can recognize common gestures based on Fresnel zone mode and decision tree classifier algorithm. It also achieves complex and successive gesture recognition based on the hidden Markov model.
- We design and implement the prototype system based on commodity WiFi cards. We take comprehensive experimental evaluation in different indoor environments. Experimental results show that iGest achieves strong robustness and high accuracy of 91% on average.

The structures in the rest of the paper are as follows: In Section 2, we discuss the advantages and disadvantages of gesture recognition in different ways. Our preliminary work is described in Section 3. In the fourth section, we propose a gesture recognition system and discuss the key parts of the system in detail. In addition, we describe the experimental and performance analysis results in section 5 and analyze the experimental results in section 6.

II. RELATED WORK

In general, the gesture recognition studies can be divided into two types: contact gesture recognition and contactless gesture recognition. The contact gesture recognition technique use the wearable devices to sense the movement of the arms. By contrast, the contactless gesture recognition is more convenient to users. With the development of ubiquitous computing, the contactless gesture recognition is paid more attention to.

A. CONTRACT GESTURE RECOGNITION

The earliest implementations of gesture recognition use special wearable devices, such as smart wristband [12]. The wristband uses a 9-axis inertial measurement unit to capture changes in the direction of the arm. The start and end of smoking can be judged by gestures, so smoking gestures and conversations can be detected in real-time. Aiming at the multi-sensory glove recognition system for paralyzed hand gestures, Nelson *et al.* [13] design an EOG-based headband

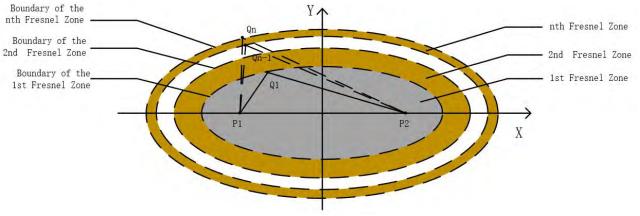


FIGURE 2. Geometry of the Fresnel zone.

using an accelerometer and flex sensors to detect eye and hand gestures. The headband uses a glove and textile electrodes, so the users can control the home's smart appliances by using EOG-based headband. Xu *et al.* [14] use the basic features of gyroscope and accelerometer data to construct a classifier which can uniquely recognize gestures. Further, for expanding gesture recognition, finger writing can also be accurately recognized.

B. CONTRACTLESS GESTURE RECOGNITION

In addition, some studies also use camera devices for gesture recognition, such as the imaging technology of the depth camera [15], multi-PC/ camera system [1]. The ellipsoid is used to fit the contour data of the human body in the image captured by the camera. The system realizes the real-time 3D reconstruction of the moving human body. And ASL [16] measured static postures and hand motions by an optical camera, and the pseudo-two-dimensional hidden Markov model could classify them.

In recent years, most researchers try to use wireless sensing to realize gesture recognition. Bo Chen *et al.* [17] implement their receiver system on the NI-based SDR platform. With full-training, the system can detect a keystroke within one key offset. Wang *et al.* [18] transform the data matrix into 2-dimensional radio images and constructs image-based feature maps for the input of DNN. Due to the CNN method extracts features from local fields, it can reduce the complexity of training and reduce costs. It is necessary for users to do a great deal of training and learning in different positions to use these methods. Then the collected gesture signals are compared with the training generated library and recognized finally.

With the development of WiFi devices, there are many researchers utilize channel state information (CSI) to achieve gesture recognition. Wang *et al.* [8] develop an activity recognition system that uses CSI to identify human activities under WIFI equipment. E-eyes [19] utilizes CSI values for recognizing household activities. In addition, fine-grained gesture recognition systems based on Wi-Fi CSI were proposed [20]–[22].

Compared with the existing methods, this paper proposes a novel method of gesture recognition to overcome the dependence of gestures recognition on the training learning. We achieve an intelligent and robust contactless gesture recognition based on a Fresnel zone model where the gesture movement is detected by a modeling method.

III. EXPERIMENTAL OBSERVATIONS

In this part, we will analyze the principle. The theory of Fresnel zone and the effects of gestures in the Fresnel zone model are briefly introduced. Then, according to the observed experimental phenomena, the causes of the effects of gestures in the Fresnel zone model are analyzed.

A. PRINCIPLE ANALYSIS

The channel state information (CSI) estimates the channel information by representing the channel attributes of the communication link [23]. The channel information is fine-grained and contains detailed phase and amplitude information about each subcarrier. The physical layer CSI can describe the propagation path of each subcarrier and can reduce the multipath effect. Wi-Fi 802.11n specification utilizes an OFDM based transmission scheme. The entire bandwidth is divided into multiple subcarriers. 20 MHz channels with different frequencies have 56 OFDM subcarriers (indexes from -28 to -1 and 1 to 28) with a carrier interval of 0.3125 MHz. The total bandwidth occupied is 17.8 MHz.

Fresnel zone is a series of concentric ellipsoids with alternating intensity caused by the propagation of light or radio waves in free space. There will occur constructive and destructive interference when paths of different lengths enter and leave the phase. Assume that P1 and P2 are two radio transceivers, as shown in Fig 2. Fresnel zone model is a concentric ellipsoid with a focus in a pair of transceivers. For a given radio wavelength of λ, n ellipses contained in the Fresnel zone can be constructed by ensuring:

$$|P_1Q_n| + |Q_nP_2| - |P_1P_2| = n\lambda/2 \tag{1}$$

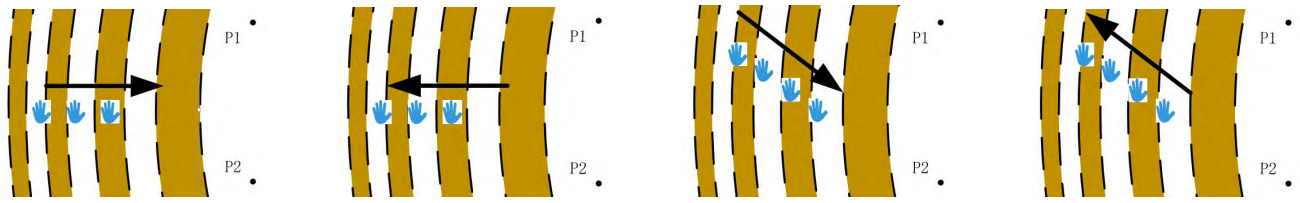
where Q_n is a point in the *nth* ellipse. The innermost ellipse is defined as the first Fresnel Zone. The elliptical ring between the first ellipse and the second ellipse is defined as the second Fresnel zone. The *nth* Fresnel zone corresponds to the elliptic ring between the $n-1$ ellipse and the *nth* ellipse.

B. OBSERVATION RESULTS

When the users make gestures in the Fresnel zone, radio signals are considered to be transmitted from the transmitter to the receiver via two paths. One path is direct transmission (line-of-sight path). The other path is reflected by the human (reflection path) [24], [25]. The signal of LOS remains unchanged, and only the reflected signal of the palm surface continuously changes. The two signals are combined to produce a superposition signal at the receiving end.

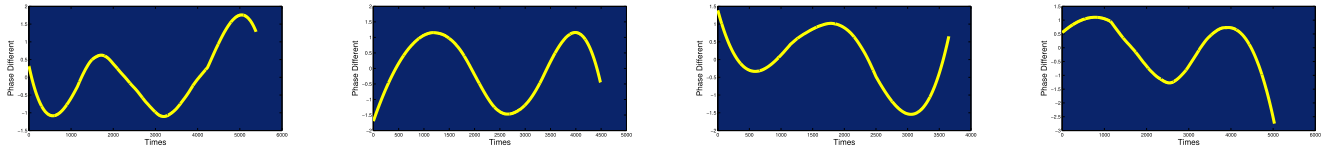
When the signal is reflected, the phase of the signal reverses and changes by π. If the phase difference between the two signals is 2π, they would enhance the signal strength. If the phase difference between the two signals is π, each signal cancels each other, and reduces the signal strength. As a result, when a gesture crosses the Fresnel boundary, the received signal will have a peak or a valley.

In practice, we construct extensive experiments to understanding above analysis. The experimental results are shown



(a) Push gesture diagram (b) Pull gesture diagram (c) Right gesture diagram (d) Left gesture diagram

FIGURE 3. Illustration of gestures movement model in the Fresnel zone.



(a) Push gesture (b) Pull gesture (c) Right gesture (d) Left gesture

FIGURE 4. Phase difference waveform with different gestures.

in the Fig. 4. From the figures, we can obtain the following insights:

- Considering that when the user make gestures in the Fresnel zone, the received signal will appear as a peak or valley as the palm crosses the each boundary of Fresnel zone. When the palm moves along the ellipse, the received signal remains stable as the length of the reflected signal path remains the same. When the palm continuously spans multiple Fresnel zone boundaries, the received signal looks like a sinusoidal waveform, and the peaks of this sinusoidal waveform correspond to the boundaries of the odd/even Fresnel zone, respectively.
- As shown in Fig. 3(a) and Fig. 3(b), when the user gestures at a constant speed in Fresnel zone, the gestures of push and pull are vertically cut across the Fresnel zone boundaries. It takes shorter distance to cross two Fresnel zone boundaries. So the time from the peak to the adjacent valley is less. If the user makes a push gesture, the palm moves closer to the center of the Fresnel zone. As it approaches the center of Fresnel, the distance between the two Fresnel zone boundaries may increase. As shown in Fig. 4(a), the absolute values of wave peaks and valleys with phase difference in the Fresnel zone tend to increase. And while the user makes a pull gesture that palm moves away from the center of Fresnel zone, it's the opposite to the gesture of push. As shown in Fig. 4(b), the absolute values of wave peaks and valleys with phase difference in Fresnel zone may decrease.
- As shown in Fig. 3(c) and Fig. 3(d), the user gestures at a constant speed in the Fresnel zone, the gestures of right and left are slanted to cut the Fresnel zone boundaries. It takes longer distance to cross two boundaries. The time of phase difference waveform from peak to adjacent valley is obviously more than a push or pull gesture. If the user makes a right gesture, the palm moves close to the center of the Fresnel zone. As shown in Fig. 4(c), the absolute values of wave peaks and valleys with phase difference in Fresnel zone tend to increase. As shown in Fig. 4(d), when the user makes a left gesture, the time of phase difference waveform from peak to adjacent valley is obviously longer, and the absolute values of wave peaks and valleys with phase difference in Fresnel zone is decreasing.
- When users do the gesture movement to cut across the Fresnel zone boundaries. The closer the angle of motion is to vertical cutting, the less time it takes to cross the Fresnel zone boundaries. The phase difference waveform of the Fresnel zone has less time from peak to the adjacent valley. When the motion is close to the center of Fresnel, the absolute value of the waveform peaks and valleys of the phase difference increases. On the contrary, it shows a decreasing trend. From these, we can judge the gestures made by the human body according to the trend of the movement.

IV. iGest DESIGN

In this part, we design a robust contactless gesture recognition system based on above observations [26]. As shown in the Fig. 5, the system is divided into three modules. The first is the data preprocessing module. In this module, raw gesture data is processed through de-noised in the time windows. The second module is the feature extraction. In this module, signals features in the sliding window are extracted. After extracting the phase difference between two subcarriers, the peak and valley characteristics of the phase difference are analyzed. Then we encode the eigenvalues according to the conclusion of the analysis. The last one is the activity analysis module. The gesture actions are classified based on received feature codes. Finally, a hidden markov model is applied in classification and accuracy judgment.

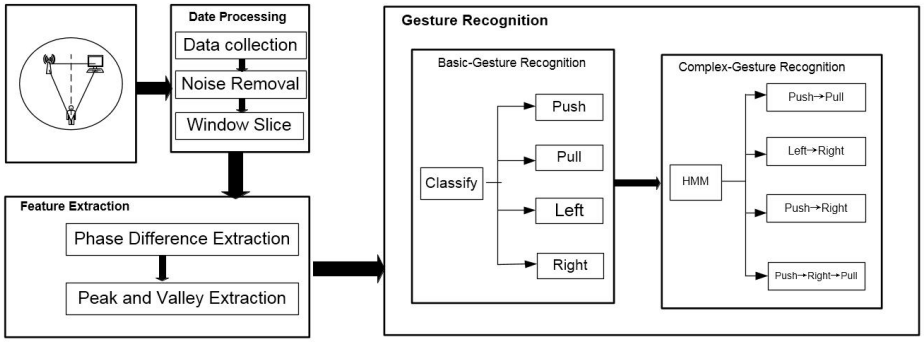


FIGURE 5. The iGest system architecture.

A. DATA PROCESSING

1) DATA COLLECTION

The CSIs from the WiFi communication are collected at receivers in the form of real-time streams. Then the data is sent to a server to process [27], [28], [42]. The amplitude and phase of CSI are sent every few minutes according to the synchronization preamble.

2) NOISE REMOVAL

The noise signal is smoothed by filtering technique [29], [30]. There is a lot of noise interference in the CSI data collected. It is sensitive to noise when the system using cross-correlation theory to measure two OFDM subcarriers. The data filtering of each subcarrier is required to keep the same phase. We use the Savitzky-Golay filter to smooth the signal to solve the above problem [31].

3) WINDOW SLICE

The time window is used to divide the temporal signal features into independent slices for recognizing different gestures. When a user makes gestures indoors, the speed of gestures in a short window can be regarded as constant. So the choice of window size is particularly important. If the window size is too small to cover the entire cycle, it cannot be reliable to estimate. On the contrary, if the size of the window is too large, users may gesture at different speeds in the window, and it will cause inaccurate estimates.

In practice, the speed of normal people gesture in the indoor environment is about 0.3 to 2 meters per second. In the Fresnel zone, the peak distance is slightly greater than $\lambda/2$, which is about 3cm in the 5GHz band. As a result, CSI power fluctuates about (0.3,2) / 0.03 times per second, equivalent to 10 Hz to 70Hz. A 0.1-second window contains 1 to 7 periods. Therefore, we think 0.1 second as window size is better, which is evaluated in our experiment evaluation.

B. FEATURE EXTRACTION

1) PHASE DIFFERENCE EXTRACTION

Once two subcarriers are selected in the phase of phase difference extraction, a longer reflection path will result in a larger phase difference $\Delta\rho$. When the position of the user is fixed, the greater the frequency difference between the two subcarriers is, the larger the phase difference $\Delta\rho$ is. If the difference between the two subcarriers is too small, the two waveforms are too close to be differential. But the large difference between the two subcarriers will cause phase ambiguity [32]. Therefore, the selection of subcarriers is particularly important.

Because multipath can lead to waveform distortion and random phase shift. So we select $\pi/2$ as the maximum allowable phase delay to guide the selection of subcarriers. Mathematically, we can calculate the frequency difference between the subcarriers of wavelength λ_1 and λ_2 according to the phase difference $\Delta\rho$:

$$\Delta f = \frac{c \cdot \frac{\pi}{2}}{2\pi \cdot (d_n - d_0)} \tag{2}$$

According to the equation :

$$\Delta f = \frac{c}{4n\lambda} \tag{3}$$

If our room size is 6*6 meters and the distance between transmitter and receiver is 4 meters. The maximum possible length of a single reflection path is less than 15 meters. When $\Delta\rho$ is limited to $\pi/2$, the maximum allowable frequency-frequency difference is 6.8MHz. If the WiFi card is configured with 40 MHz bandwidth, the CSI value from adjacent OFDM subcarriers is 1.25MHz according to the 802.11n-2009 specification. Therefore, we select two CSI subcarriers, and they are interred per five indicators, such as 1 and 6, 2 and 7, etc.

2) PEAK AND VALLEY EXTRACTION

By analyzing the time delay distribution, the delay between two subcarrier windows is calculated. The selection of eigenvalues is estimated by analyzing the delay distribution.

We respectively calculate the spatial feature α and temporal feature β between adjacent peak or adjacent valley as the feature of extraction and encode the extracted eigenvalues. If the phase difference of the first peak is larger than that of the second peak, the α is defined as 1. Conversely, the α is set to 0. If the time difference between two peaks or valleys is

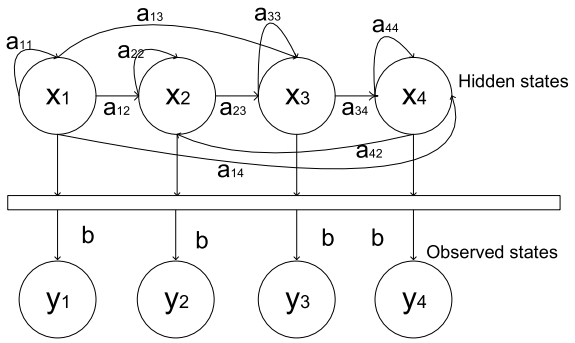


FIGURE 6. Illustration of the HMM model.

larger than the threshold (1500 ms in our experiments), the β is defined as 1, or the β is defined as 0. In the following, we use the two features α and β to classify the gestures.

C. GESTURE RECOGNITION

1) BASIC-GESTURE RECOGNITION

In the section, we introduce the scheme of gesture recognition. First, a decision tree classifier is used to deal with the collected features to improve the accuracy of detection [33]–[36]. Then, the gesture action is judged according to the extracted features.

In the training stage, we construct extensive experiments to select the training data. We respectively calculate the α and the β values in the time windows for each gestures. Then we make a training model for the gesture recognition. Moreover, because the α and the β are independent of the users' positions, we pay no attention to the locations and environments during the training model. Hence, it needs no retraining when the iGest system is deployed in a new environment.

2) COMPLEX-GESTURE RECOGNITION

In real life, many gestures are composed of the basic actions. For example, the gesture of refusal "No" is a combination of gesture "left" and "right", the calling gesture "Come" is a combination of gesture "push" and "pull". We detect the movement trend in the time window and infer his next action trend to complete the complex gesture recognition.

The hidden markov model is applied to infer the next gesture [37]. In our system, dividing the collected data sets into two parts, 70% for generating conversion probability, and 30% for testing validation.

We can define the HMM model as $\lambda = \{\pi, A, B\}$, where A is a transition probability matrix of hidden states, B is a obfuscation matrix of observed states and π is an initial probability distribution. Then we use a Viterbi algorithm [**?**] to infer the hidden state according to the observed state, that is, gesture movements can be inferred from our detection data.

We define that A is a transition probability matrix of hidden states, a hidden state is represented by S. The S_j is the hidden state at $t+1$ time and the state S_i is at t time. N is the number of hidden states, that is, the number of gestures.

$$A_{ij} = P(S_j|S_i), \quad 1 \le i, j \le N \tag{4}$$

B is defined as the obfuscation matrix. The observed state is represented by O. In fact, B denotes the probability of the observed state O_i when the hidden state S_j at the t moment. M is defined as the number of observed states, that is, the number of features observed in the experiments.

$$B_{ij} = P\left(O_i|S_j\right), \quad 1 \le i \le M, \ 1 \le j \le N \tag{5}$$

π is an initial probability distribution and it is a one-dimensional matrix.

$$\pi = [\, P(S_1) \quad P(S_2) \quad \cdots \quad P(S_N) \,] \tag{6}$$

Using recursive methods to reduce computational complexity is the core principle of Viterbi algorithm. Knowing that a gesture is a state. The relevant values are defined as follows: The probability of reaching an intermediate state is δ. Hidden state (previous gesture) is x. Observable output (current gesture) is y. The state transition probability (the probability that x becomes y) is a. Output probability (probability of gesture) is b, as shown in Fig. 6.

When $t = 1$, the relationship between probability of the i^{th} state and the transition probability of the ith state to the observable sequence k_1 is as follows:

$$\delta_1(i) = \pi(i)\, b_{ik_1} \tag{7}$$

Calculate the partial probability of $t > 1$ time, and the probability that x occurs in the end of a sequence depends on previous state before it.

Based on these three probability values (Maximum probability of the previous state, Conversion probability between two states, Output probability of the next state), we can get the maximum probability of the next state:

$$\Pr(X_t) = \max_{i=A,B,C} \Pr(i_{t-1}) \times \Pr(X|i) \times \Pr(observation_t|X) \tag{8}$$

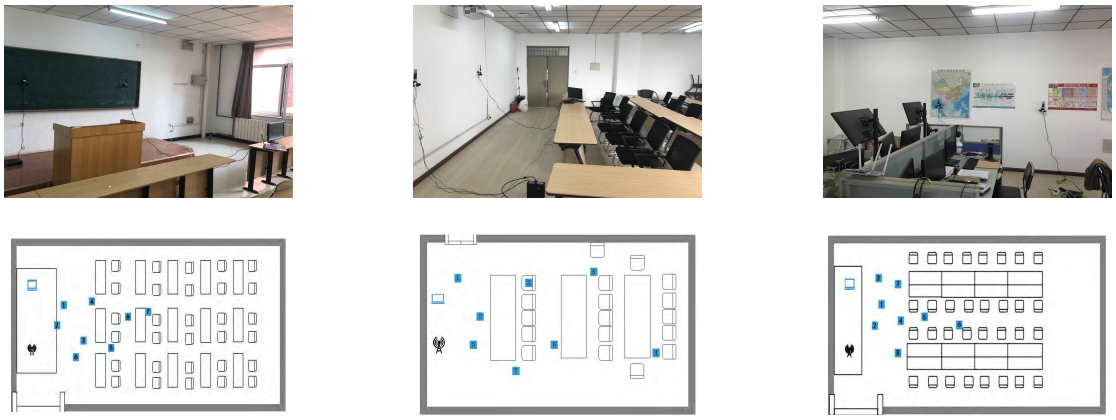
By extending the above expression, we can get the formula for calculating the maximum partial probability of the ith state of the observable state at t time:

$$\delta(i) = \max_j(\delta_{t-1}(j)a_{ji}b_{ik_t}) \tag{9}$$

where it represents the probability of transition from state j to state i, and b_{ik_t} represents the probability that state i is observed as k_t.

There is a partial optimal probability of $\delta(i,t)$ in each intermediate state and final state. But our goal is to find the hidden state sequence with the highest probability.

To calculate the partial probability at time t, we only need to know the partial probability at time $t-1$. So we only need to record the state that caused the maximum partial probability at time t.



(a) Classroom (b) Meeting room (c) Laboratory room

FIGURE 7. Three different environments tested in our experiments.

The previous state of the maximum local probability of a state can be recorded with a backward pointer φ:

$$\varphi_t(i) = \arg\max_j(\delta_{t-1}(j)a_{ji}) \tag{10}$$

At this point, the current gesture can be judged based on the acquired data and the previous gesture can be inferred. The complex gesture is judged according to the combination of the two gestures.

V. EXPERIMENTAL EVALUATION

In this section, we evaluate the performance of our contactless gesture recognition system. First, we describe the experimental setup and experimental environment. Then the detailed experimental results are given and analyzed comprehensively. Finally, we present the discussion about our techniques.

A. EXPERIMENTAL SETUP

We designed experiments in three test environments [10], [38], [39], as shown in Fig. 7.

1) **Testbed-1:** We constructed our experiments in a classroom room, which was 6m×6m. It was surrounded by a large number of chairs, desks. One pair of TX and RX was placed apart about 3m, about 1.4m above the floor and we chose 8 reference test locations according to the floor flan in Figure 7(a).
2) **Testbed-2:** We constructed our experiments in a meeting room, which was 4m×5m. It was surrounded by a large number of chairs, desks. One pair of TX and RX was placed apart 2.5m, about 1.4m above the floor and we chose 8 reference test locations according to the floor flan in Figure 7(b).
3) **Testbed-3:** We tested iGest in a typical laboratory room, which was over approximately $12m^2$. It was surrounded by a large number of chairs, desks and computers. One pair of TX and RX was placed apart 2m, about 1.4m above the floor and we chose 8 reference test locations according to the floor flan in Figure 7(c).

In our experiments, we used two mini host PCs (Intel® Desktop Board D2700MUD, 1GB RAM, Mini-ITX SIZE 170 *mm* × 170 *mm*, three external antennas) equipped with Intel 5300 NIC and running 32-bit Ubuntu Linux (version 10.04LTS of the Server Edition) were used as the receiver (RX) and the transmitter (TX). The *iwlwifi* firmware is modified as [40], [41] to start traffic flow and export CSI of each packet, i.e., a group of 30 CSI for further analysis. Each wireless card is equipped with a full outgoing antenna. We only select one antenna to receive or send packets. The transmitter is configured to send packets in injection mode. The transmitter drops some packets in a predefined mode every 10 seconds as a synchronization signal. Therefore, two receivers can align data based on this signal. To capture finer Fresnel phase differences between subcarriers, gestures can be identified more accurately. We chose the 5.32 GHz band to experiment with a bandwidth of 40 MHz. To capture the signal fluctuations generated by human gesture recognition better, we set the sampling frequency to 500 packets per second.

We employed four volunteers during the experiments. First, each volunteer was assigned four simple gestures: push, pull, left, and right at a given test point (8 test points). Each group of gestures was repeated 5 times. There were $4 * 8 * 5 = 160$ groups of experiments in each environment. And then each volunteer was also assigned four complex gestures: push->pull, left->right, push->right and push->right->pull.

B. RESULT EVALUATION

We take extensive experimental evaluation in the test environments to test the performance of iGest systems.

1) OVERALL PERFORMANCE OF GESTURE RECOGNITION

Fig. 8 shows the confusion matrix for four basic gestures recognition performance. Each row represents the actual gestures performed by the user, and each column represents the

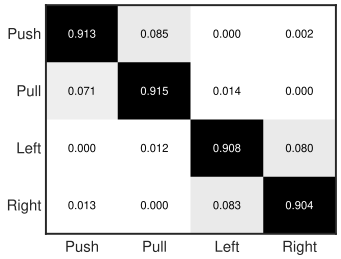


FIGURE 8. Accuracy of basic gestures recognition.

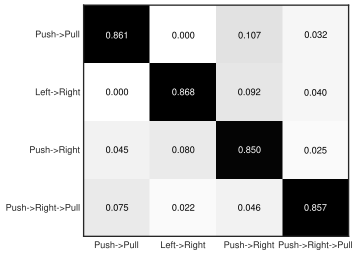


FIGURE 9. Accuracy of complex gestures recognition.

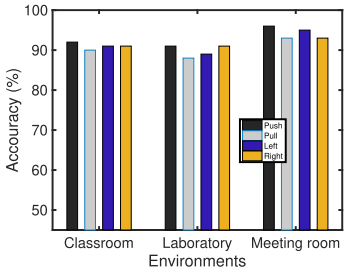


FIGURE 10. Performance of iGest in different environments.

classified gestures. Each element in the matrix corresponds to the fraction of the gesture in the row that is classified as the gesture in the column.

When classifying the four basic gestures, the average accuracy is 91% on average. Compared with the accuracy of four gestures recognition, it is found that the recognition accuracy is similar. Among them, the recognition accuracy of push and pull is relatively high, while the accuracy of left and right is relatively low. It shows that it is easier to identify the gestures that vertically go through Fresnel zone, by compared with the gestures that horizontally cut Fresnel zone.

Moreover, we evaluate the performance of iGest on identifying the complex gestures. In our experiments, we test four complex gestures when the volunteers stand at different locations. As shown in the Figure 9, the accuracy of complex gesture recognition is above 85% on average.

2) ACCURACY IN DIFFERENT SCENARIOS

Fig. 10 shows the average accuracy of detection when the volunteers takes the experiments in different experimental environments. The accuracy of gestures recognition is 91.7%

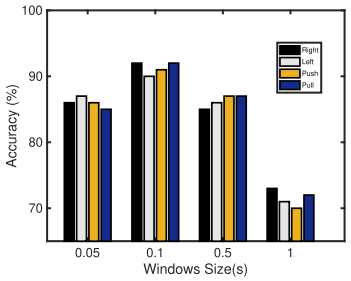


FIGURE 11. Performance of iGest with different window sizes.

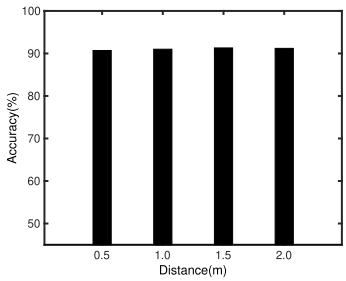


FIGURE 12. Performance of iGest with different distances to LOS path.

on average when the volunteers were in the classroom, and 90.2% on average in the laboratory, 91.1% on average in the meeting room. It can be seen that the accuracy of the iGest system does not change dramatically among different experimental environments. The detection accuracy is stable about 91% on average, so the iGest system is relatively robust for the environments.

3) ACCURACY WITH DIFFERENT TIME WINDOW SIZES

The sliding window size may affect the overall performance of the system. If the window size is too small to cover the entire cycle, the accuracy will be reduced; On the contrary, if the window size is too large, people may make gestures at different speeds in this window, and the accuracy of detection may be also reduced.

From Fig. 11, it is observed that when we set the sliding window size as 0.1 second, the detection performance of the system is the highest with the accuracy up to 91%. Therefore, the sliding window size of 0.1s is the best choice for the iGest system.

4) ACCURACY WITH DIFFERENT DISTANCES TO LOS PATH

Fig. 12 shows the average detection accuracy when the users stand at different distances to the link of TX-RX pair. When the distance is very small, the accuracy of gesture detection is about 90.7% on average. The accuracy is about 91.3% in the moderate distance, and 91% in far distance. In our system, though the distances between the volunteer and the link of TX-RX pair are different, the accuracy of iGest system is similar about 91% on average. Hence, the distance of the user's position to the center of the Fresnel zone model has little effect on the iGest system in a specified range.

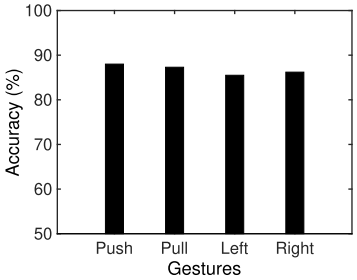


FIGURE 13. Performance of iGest in the NLOS environment.

5) PERFORMANCE OF iGest IN NLOS ENVIRONMENTS

Sometimes, the users may locate in a non line-of-sight (NLOS) environment, such as behind the wall or behind a door. Hence, the performance of iGest in NLOS environments is needed to be evaluated. In our experiments, the volunteer stands behind the door and makes four gestures. Fig. 13 shows that the average detection accuracy of four gestures is 86.7% when the user locates in the NLOS environment. It is also observed that he detection accuracy of push or pull gestures is higher than that of left and right gestures.

VI. CONCLUSIONS

This article demonstrates the ability and robustness to recognize human gesture information by a ready-made WiFi device based on Fresnel phase analysis theory. We conduct a comprehensive theoretical study and take deeply experimental observations. Furthermore, we design an intelligent gesture recognition system (iGest) using the commercial WiFi devices. Our experimental results show that the iGest can recognize gestures in different indoor environments, and the overall recognition rate error is less than 10%. The results promote the establishment of our cost-effective gesture recognition system based on channel state information of WiFi. We believe that our system can be applied to a wider range of macro-micro human interaction applications.

REFERENCES

[1] G. K. M. Cheung, T. Kanade, J.-Y. Bouguet, and M. Holler, "A real time system for robust 3D voxel reconstruction of human motions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 2000, pp. 714–720.

[2] J. M. Rehg and T. Kanade, "Visual tracking of high DOF articulated structures: An application to human hand tracking," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 1994, pp. 35–46.

[3] D. Kim, O. Hilliges, S. Izadi, A. D. Butler, J. Chen, I. Oikonomidis, and P. Olivier, "Digits: Freehand 3D interactions anywhere using a wrist-worn gloveless sensor," in *Proc. 25th Annu. ACM Symp. User Interface Softw. Technol.*, Oct. 2012, pp. 167–176.

[4] T. Park, J. Lee, I. Hwang, C. Yoo, L. Nachman, and J. Song, "E-gesture: A collaborative architecture for energy-efficient gesture recognition with hand-worn sensor and mobile devices," in *Proc. 9th ACM Conf. Embedded Netw. Sensor Syst.*, Nov. 2011, pp. 260–273.

[5] C. Xiang, P. Yang, C. Tian, L. Zhang, H. Lin, F. Xiao, M. Zhang, and Y. Liu, "CARM: Crowd-sensing accurate outdoor RSS maps with error-prone smartphone measurements," *IEEE Trans. Mobile Comput.*, vol. 15, no. 11, pp. 2669–2681, Nov. 2016.

[6] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proc. 19th Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2013, pp. 27–38.

[7] F. Adib and D. Katabi, "See through walls with WiFi!" in *Proc. ACM SIGCOMM Comput. Commun. Rev.*, vol. 43, no. 4, pp. 1–64, 2013.

[8] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of WiFi signal based human activity recognition," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2015, pp. 65–76.

[9] J. Yang, Y. Ge, H. Xiong, Y. Chen, and H. Liu, "Performing joint learning for passive intrusion detection in pervasive wireless environments," in *Proc. INFOCOM*, Mar. 2010, pp. 1–9.

[10] H. Zhu, F. Xiao, L. Sun, R. Wang, and P. Yang, "R-TTWD: Robust device-free through-the-wall detection of moving human with WiFi," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 5, pp. 1090–1103, May 2017.

[11] G. Gui, H. Huang, Y. Song, and H. Sari, "Deep learning for an effective nonorthogonal multiple access scheme," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8440–8450, Sep. 2018.

[12] A. Parate, M.-C. Chiu, C. Chadowitz, D. Ganesan, and E. Kalogerakis, "RisQ: Recognizing smoking gestures with inertial sensors on a wristband," in *Proc. 12th Annu. Int. Conf. Mobile Syst., Appl., Services*, Jun. 2014, pp. 149–161.

[13] A. Nelson, J. Schmandt, P. Shyamkumar, W. Wilkins, D. Lachut, N. Banerjee, S. Rollins, J. Parkerson, and V. Varadan, "Wearable multi-sensor gesture recognition for paralysis patients," in *Proc. SENSORS*, Nov. 2013, pp. 1–4.

[14] C. Xu, P. H. Pathak, and P. Mohapatra, "Finger-writing with smartwatch: A case for finger and hand gesture recognition using smartwatch," in *Proc. 16th Int. Workshop Mobile Comput. Syst. Appl.*, Feb. 2015, pp. 9–14.

[15] S. Agrawal, I. Constandache, S. Gaonkar, R. R. Choudhury, K. Caves, and F. DeRuyter, "Using mobile phones to write in air," in *Proc. 9th Int. Conf. Mobile Syst., Appl., Services*, Jun. 2011, pp. 15–28.

[16] T. Starner and A. Pentland, "Real-time American sign language recognition from video using hidden Markov models," in *Proc. Int. Symp. Comput. Vis. (ISCV)*, Nov. 1997, pp. 227–243.

[17] B. Chen, V. Yenamandra, and K. Srinivasan, "Tracking keystrokes using wireless signals," in *Proc. 13th Annu. Int. Conf. Mobile Syst., Appl., Services*, May 2015, pp. 31–44.

[18] Q. Gao, J. Wang, X. Ma, X. Feng, and H. Wang, "CSI-based device-free wireless localization and activity recognition using radio image features," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 10346–10356, Nov. 2017.

[19] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu, "E-eyes: Device-free location-oriented activity identification using fine-grained WiFi signatures," in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2014, pp. 617–628.

[20] S. Sigg, S. Shi, and Y. Ji, "RF-based device-free recognition of simultaneously conducted activities," in *Proc. ACM Conf. Pervasive Ubiquitous Comput. Adjunct Publication*, Sep. 2013, pp. 531–540.

[21] W. He, K. Wu, Y. Zou, and Z. Ming, "WiG: WiFi-based gesture recognition system," in *Proc. 24th Int. Conf. Comput. Commun. Netw. (ICCCN)*, Aug. 2015, pp. 1–7.

[22] M. A. A. Al-Qaness and F. Li, "WiGeR: WiFi-based gesture recognition system," *Int. J. Geo-Inf.*, vol. 5, no. 6, p. 92, 2016.

[23] H. Huang, Y. Song, J. Yang, G. Gui, and F. Adachi, "Deep-learning-based millimeter-wave massive MIMO for hybrid precoding," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 3027–3032, Mar. 2019.

[24] Z. Yang, P. H. Pathak, Y. Zeng, X. Liran, and P. Mohapatra, "Vital sign and sleep monitoring using millimeter wave," *ACM Trans. Sensor Netw.*, vol. 13, no. 2, p. 14, Jun. 2017.

[25] Y. Wang, M. Liu, J. Yang, and G. Gui, "Data-driven deep learning for automatic modulation recognition in cognitive radios," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 4074–4077, Apr. 2019.

[26] S. Ren, H. Wang, B. Li, L. Gong, H. Yang, C. Xiang, and B. Li, "Robust contactless gesture recognition using commodity WiFi," in *Proc. EWSN*, Feb. 2019, pp. 273–275.

[27] X. Zhang, Z. Yang, Y. Liu, J. Li, and Z. Ming, "Toward efficient mechanisms for mobile crowdsensing," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1760–1771, Feb. 2017.

[28] X. Zhang, L. Liang, C. Luo, and L. Cheng, "Privacy-preserving incentive mechanisms for mobile crowdsensing," *IEEE Pervasive Comput.*, vol. 17, no. 3, pp. 47–57, Jul./Sep. 2018.

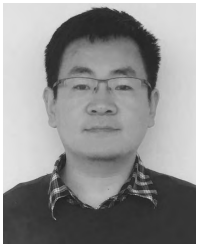
[29] Z. Yang, L. Jian, C. Wu, and Y. Liu, "Beyond triangle inequality: Sifting noisy and outlier distance measurements for localization," *ACM Trans. Sensor Netw.*, vol. 9, no. 2, p. 26, Mar. 2013.

[30] B. Sun, Q. Ma, S. Zhang, K. Liu, and Y. Liu, "iSelf: Towards cold-start emotion labeling using transfer learning with smartphones," *ACM Trans. Sensor Netw.*, vol. 13, no. 4, p. 30, Dec. 2017.

[31] S. Tan and J. Yang, "WiFinger: Leveraging commodity WiFi for fine-grained finger gesture recognition," in *Proc. 17th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, Jul. 2016, pp. 201–210.
[32] X. Zhang, Z. Yang, and Y. Liu, "Vehicle-based bi-objective crowdsourcing," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 10, pp. 3420–3428, Oct. 2018.
[33] C. Chen, S. Jiao, S. Zhang, W. Liu, L. Feng, and Y. Wang, "TripImputor: Real-time imputing taxi trip purpose leveraging multi-sourced urban data," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 10, pp. 3292–3304, Oct. 2018.
[34] Z. Li, Y. Zhang, and Y. Liu, "Towards a full-stack devops environment (platform-as-a-service) for cloud-hosted applications," *Tsinghua Sci. Technol.*, vol. 22, no. 1, pp. 1–9, Feb. 2017.
[35] C. Chen, D. Zhang, X. Ma, B. Guo, L. Wang, Y. Wang, and E. Sha, "Crowddeliver: Planning city-wide package delivery paths leveraging the crowd of taxis," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 6, pp. 1478–1496, Jun. 2017.
[36] L. Gong, Y. Zhao, X. Chaocan, Z. Li, C. Qian, and P. Yang, "Robust lightweight magnetic-based door event detection with smartphones," *IEEE Trans. Mobile Comput.*, to be published.
[37] M. R. Hassan and B. Nath, "Stock market forecasting using hidden Markov model: A new approach," in *Proc. 5th Intell. Syst. Design Appl. (ISDA)*, Sep. 2005, pp. 192–196.
[38] H. Wang, D. Zhang, K. Niu, Q. Lv, Y. Liu, D. Wu, R. Gao, and B. Xie, "MFDL: A multicarrier Fresnel penetration model based device-free localization system leveraging commodity Wi-Fi cards," 2017, *arXiv:1707.07514*. [Online]. Available: https://arxiv.org/abs/1707.07514
[39] B. Xie, K. Chen, G. Tan, M. Lu, Y. Liu, J. Wu, and T. He, "LIPS: A light intensity–based positioning system for indoor environments," *ACM Trans. Sensor Netw.*, vol. 12, no. 4, p. 28, Nov. 2016.
[40] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 1, p. 53, Jan. 2011.
[41] Z. Li, W. Wang, C. Wilson, J. Chen, C. Qian, T. Jung, L. Zhang, K. Liu, X. Li, and Y. Liu, "FBS-radar: Uncovering fake base stations at scale in the wild," in *Proc. NDSS*, Mar. 2017, pp. 1–15.
[42] Y. Zheng, Z. Yang, J. Yin, C. Wu, K. Qian, F. Xiao, and Y. Liu, "Combating cross-technology interference for robust wireless sensing with COTS WiFi," in *Proc. 27th Int. Conf. Comput. Commun. Netw. (ICCCN)*, Jul. 2018, pp. 1–9.

SHUJIE REN received the B.S. degree in communication engineering from the Tianjin University of Technology, China, in 2017, where she is currently pursuing the M.Sc. degree in computer technology. Her main research interests include the architecture of wireless security, indoor locations, and wireless perception.

HUAIBIN WANG received the B.Sc. degree in computer and application from the Jilin University of Technology, China, in 1982, and the M.Sc. degree in computer science and technology from the Harbin Institute of Technology, China, in 1988. He is currently a Professor with the Tianjin University of Technology. His current research interests include network information security, computer simulation technology, and computer software and its application.



LIANGYI GONG received the B.E. and Ph.D. degrees from Harbin Engineering University, China, in 2010 and 2016, respectively. From 2016 to 2019, he was an Assistant Professor with the School of Computer Science and Engineering, Tianjin University of Technology, China. He is currently a Postdoctoral Researcher with the School of Software and BNRist, Tsinghua University. His major research interests include mobile/pervasive computing and network security.

CHAOCAN XIANG received the B.S. and Ph.D. degrees in computer science and engineering from the Institute of Communication Engineering, PLA University of Science and Technology, China, in 2009 and 2014, respectively. He is currently an Associate Professor with Chongqing University, China. His current research interests include mobile/pervasive computing, crowd-sensing networks, and the IoT.

XUANGOU WU received the Ph.D. degree from the School of Computer Science and Technology, University of Science and Technology of China, Hefei, China, in 2013. He is currently an Associate Professor with the School of Computer Science and Technology, University of Anhui Technology, Maanshan, China. His research interests include wireless sensor networks, crowdsourcing, privacy protection, and datacenter networks.

YUFENG DU received the B.S. degree in information and computing science from Shanxi Agricultural University, in 2017. He is currently pursuing the M.Sc. degree in computer technology with the Tianjin University of Technology, China. His main research interests include the architecture of wireless security, indoor locations, and wireless perception.